

STRUCTURE AND METHOD TO ENHANCE BOTH NFET AND PFET
PERFORMANCE USING DIFFERENT KINDS OF STRESSED LAYERS

DESCRIPTION

BACKGROUND OF THE INVENTION

5 Field of the Invention

The present invention generally relates to the
manufacture of transistors for integrated circuits
and, more particularly, to the production of
complementary pairs of field effect transistors of
enhanced performance at extremely small scale.

Description of the Prior Art

Performance and economic factors of integrated
circuit design and manufacture have caused the scale
of elements (e.g. transistors, capacitors and the
like) of integrated circuits to be drastically
reduced in size and increased in proximity on a
chip. That is, increased integration density and
proximity of elements reduces the signal propagation
path length and reduces signal propagation time and
susceptibility to noise and increase of possible
clock rates while the reduction in element size
necessary for increased integration density
increases the ratio of functionality which can be
provided on a chip to the costs of production (e.g.
wafer/chip area and process materials) per chip and,
potentially, the cost of devices containing the
chips by reducing the number of inter-chip and
inter-board connections required.

However, the immutable material properties and physical effects by which transistors and other elements function is invariably compromised as the scale of integrated circuit elements is reduced. In response, many improvements in transistor design have been made to maintain suitable levels of performance of these elements. For example, lightly doped drain (LDD) structures (now generally referred to as extension implants since heavier doping levels have been required in current minimum feature size regimes), halo implants and graded impurity profiles have been employed to counteract short channel and punch-through effects and the like, particularly in field effect transistors (FETs) which have become the active device of choice for all but the highest frequency devices. Reduction in device scale has also required operation at reduced voltages to maintain adequate performance without device damage even though operating margins may be reduced.

A principal factor in maintaining adequate performance in field effect transistors is carrier mobility which affects the amount of current or charge which may flow (as electrons or holes) in a doped semiconductor channel under control of a voltage placed on a gate electrode insulated from the channel by a very thin dielectric. Reduced carrier mobility in an FET reduces not only the switching speed/slew rate of a given transistor but also reduces the difference between "on" resistance to "off" resistance. This latter effect increases susceptibility to noise and reduces the number of and/or speed at which downstream transistor gates (capacitive loads) can be driven, sometimes referred to as fan-out. Even during the early development of

metal-oxide-semiconductor (MOS) field effect transistors and complementary MOS (CMOS) devices (widely used in integrated circuits at the present time), in particular, carrier mobility was a design concern and often required a pMOS device to be made approximately two times as large as a complementary nMOS device with which it was paired in order to obtain reasonably symmetrical operation of the CMOS pair in view of the difference in carrier mobility between electrons, the principal carrier in nMOS devices and holes, the principal carrier in pMOS devices. In more recent and critical designs, it has been shown that carrier mobility degrades in deeply scaled bulk MOS devices due to the heavy doping required to suppress short-channel effects and ultra-thin oxide effects.

It has also been shown theoretically and confirmed experimentally that mechanical stress in the channel region of an FET can increase or decrease carrier mobility significantly; depending on the sign of the stress (e.g. tensile or compressive) and the carrier type (e.g. electron or hole). Tensile stress increases electron mobility and decreases hole mobility while compressive stress increases hole mobility while decreasing electron mobility in the doped semiconductor crystal lattice forming the transistor channel. This phenomenon is well-recognized and theories concerning the physical effects by which it occurs are, in any event, unimportant to its exploitation. In this regard, numerous structures and materials have been proposed for inducing tensile or compressive force in a semiconductor material, such as shallow trench isolation (STI) structures, gate spacers, etch-stop

layers and silicide which are generally included in integrated circuit designs. It is also known to induce persistent stresses in semi-conductors for structural reasons. For example, U.S. Patents 5 6,069,049 and 6,399,976 teach that applying a film around a structure and then reducing volume of the film to compress the structure and thus prevent the propagation of defects. However, at the present state of the art, structures for producing stresses 10 can generally be made of only one type; to produce tensile stress or compressive stress but not both. Therefore, in integrated circuit designs using both pFET and nFET transistors and CMOS technology (in which the logic is implemented principally by 15 complementary pMOS and nMOS transistor pairs), in particular, an enhancement of carrier mobility in one type of transistor was necessarily accompanied by degradation of carrier mobility in the other or complementary type of transistor; yielding little, 20 if any, net performance gain, although symmetry of operation could theoretically be improved. Moreover, stress of a single type produced by such structures tends to cause warping or curling of the wafer or substrate (particularly if thin, as is a 25 current trend) which compromises later lithographic processes such as the formation of contacts and connections or, in severe or high stress cases, chip or wafer cracking; reducing manufacturing yield or (in rare cases) reliability after being put into 30 service. Further, the stress levels produced by such structures were generally difficult to control particularly since the structure dimensions are often dictated by other design concerns, such as isolation and breakdown voltages.

SUMMARY OF THE INVENTION

It is therefore an object of the present invention to provide a structure and method to enhance both nFET and pFET performance by locally and simultaneously applying tensile stress or compressive stress, respectively, through use of different kinds of stressed film.

It is another object of the present invention to provide desired stresses in transistors of an integrated circuit without causing warping or curling of a wafer or substrate.

In order to accomplish these and other objects of the invention, a method of adjusting carrier mobility in CMOS devices is provided comprising the steps of depositing a first stressed film on a wafer including a complementary pair of transistors to create a stress in the channels of the transistors, partially removing the first stressed film to relieve the stress from one channel of one transistor, and depositing a second stressed film over first and second transistors to apply a second stress to the channel of the transistor with the first stressed film removed.

Through this process, a structure that adjusts carrier mobility in CMOS transistors is provided comprising a substrate, a complementary pair of transistors each having a gate dielectric, gate electrode, and source, drain, and gate silicide regions, formed on the substrate, a first film providing tensile stress at least at the channel of one transistor, a second film providing compressive stress at least at the channel of the other

30 transistor, and a shear force isolation layer
separating the first film and the second film in at
least one area.

BRIEF DESCRIPTION OF THE DRAWINGS

5 The foregoing and other objects, aspects and
advantages will be better understood from the
following detailed description of a preferred
embodiment of the invention with reference to the
drawings, in which:

10 Fig. 1 is a cross-sectional view of the pre-
fabricated structure to which the invention is
applied in order to achieve the final exemplary pair
of nMOS and pMOS transistors shown in Fig. 6.

15 Fig. 2 is a cross-sectional view of the first
production stage of the final exemplary pair of nMOS
and pMOS transistors shown in Fig. 6 in which a
tensile nitride film is deposited in order to
produce tensile stress in the channel of devices.

20 Fig. 3 is a cross-sectional view of the second
production stage of the final exemplary pair of nMOS
and pMOS transistors shown in Fig. 6 in which oxide
is directly deposited and etched back.

25 Fig. 4 is a cross-sectional view of a third
production stage of the final exemplary pair of nMOS
and pMOS transistors shown in Fig. 6 in which some
areas of tensile nitrate are removed by method of
wet etching.

30 Fig. 5 is a cross-sectional view of a fourth
production stage of the final exemplary pair of nMOS
and pMOS transistors shown in Fig. 6 in which the
original dielectric 12 is removed and replaced with
a new dielectric layer 12 deposited on both the nMOS
and pMOS transistors, as shown on the nMOS
transistor. The nMOS transistor is then masked off
by a photo resist while the tensile film and
dielectric covering the pMOS transistor are

subjected to etching.

Fig. 6 is a cross-sectional view of a fifth production stage of the final exemplary pair of nMOS and pMOS transistors in which the photo resist is
5 etched away and compressive film is deposited thereby creating the final preferred embodiment of the invention.

Fig. 7 is a graphical depiction of stress measurements in the lateral direction as a result of
10 the exemplary configuration shown in Fig. 6.

Fig. 8a is a cross-sectional view illustrating an alternate third stage for creating a second embodiment of the invention illustrated in 8c following the completion of steps one and two (as
15 shown in Fig. 2 and Fig. 3 respectively) in which a photo-resist is applied to the nMOS transistor side of the chip or wafer.

Fig. 8b is a cross-sectional view illustrating an alternate fourth stage for creating a second
20 embodiment of the invention illustrated in Fig. 8c in which the film areas not blocked by photo-resist are removed followed by the removal of said photo-resist.

Fig. 8c is a cross-sectional view illustrating an alternate fifth stage for creating a second
25 embodiment of the invention shown here in which a compressive film is deposited across both the nMOS and pMOS transistors.

DETAILED DESCRIPTION OF A PREFERRED
EMBODIMENT OF THE INVENTION

Referring now to the drawings, and more particularly to Fig. 1, there is shown, in cross-section, an exemplary pair of complementary nMOS and pMOS transistors such as might comprise a CMOS pair in a portion of an integrated circuit. These transistors have been completed through the formation of silicide on the source, drain and gate regions to reduce resistance of contacts to be later applied thereto. The method of formation of these transistors and the particulars of their structures are unimportant to an understanding of the invention or its successful practice. It should be understood that while Fig. 1 illustrates transistors prior to the application of the invention thereto, the illustration is intended to be highly schematic for clarity and no portion of Fig. 1 is admitted to be prior art as to the invention.

In Fig. 1 and other figures, the transistors are formed on substrate 14 (which may be relatively much thicker than depicted since the transistors are possibly deeply scaled, as is the preferred application of the invention). The transistors 20 and 21 are isolated from each other by isolation oxide 16 such as shallow trench isolation which may also serve to define n-well and p-well regions of substrate 14. In this regard, transistors 20 and 21 are of complementary conductivity types and, accordingly, the portions of substrate 14 underlying these respective transistors are also oppositely doped.

The transistors 20 and 21 are otherwise of

similar construction; each having a gate dielectric 205, 215, a gate electrode 208, 218, a sidewall spacer structure 12 and source, drain and gate silicide regions as will be familiar to those
5 skilled in the art preferably formed by an implant self-aligned to sidewall spacers 11. Extension and halo implants, graded junctions and the like may also be provided.

It should be appreciated in the course of the
10 following discussion of Figs. 1-6 that the embodiment which will be described is that which is expected to be the most advantageous in most applications and integrated circuit designs, and thus preferred, as well as being an embodiment which
15 will allow the most complete appreciation of the invention to be conveyed. However, the principles of the invention described in connection with this embodiment can also be applied to provide any desired amount of stress of either sign to adjacent
20 electronic elements in any design.

Fig. 2 shows the first step taken in forming the invention shown in Fig. 6 from the basic structure shown in Fig. 1. This step includes the deposit of a stressed film 11 (e.g. tensile) over
25 the remaining structure of Fig. 1 in order to produce stress (e.g. tensile) in the channels 201, 218 of the transistors 20 and 21. Prior to the deposit of this film, the sidewall spacers 11 are optionally removed. The highly stressed film is
30 preferably silicon nitride (Si_3N_4) or silicon oxynitride ($\text{Si}_3\text{O}_x\text{N}_y$) or a combination of both. These materials can be deposited in a highly stressed form providing either tensile or compressional stress, depending on particular process parameters such as

plasma power and gas flow rates. For example, using a PECVD process, the stress level is mainly controlled by plasma power and, in general, yields a compressive stress. Thermal CVD Si_3N_4 is deposited at a temperature above 600°C and is normally tensile. Application of this tensile stress enhances the performance of the nMOS transistor 20 while decreasing the performance abilities of the pMOS transistor 21.

Fig. 3 illustrates the second step taken toward reaching the exemplary final configuration shown in Fig. 6. In this step, a dielectric 12 (e.g. oxide) is deposited onto the structure shown in Fig. 2 via high density plasma deposition (HDP) and, for example, chemical-mechanical polish, CMP stopping on nitride 11 and then etched back to oxide 12 to form the illustrated configuration. The thickness of the oxide layer 12 is not critical to the practice of the invention, although layer thickness is preferably in the range of 50-100 nm. The dielectric 12 is preferably a non-stressed oxide or other non-stressed material to serve as an etch stop for the etch back and a neutral barrier between the tensile film 11 and the later applied compressive film 13.

Fig. 4 illustrates the third step taken toward reaching the exemplary final configuration shown in Fig. 6. The dielectric 12 deposited in the previous step is now used as a mask, creating a pattern for removing areas of the stressed film 11 via wet etching. As shown in this step, the same tensile stress is still being applied to the nMOS 20 and pMOS 21 transistors increasing the performance of the nMOS transistor 20 while decreasing the

performance of the pMOS transistor 21.

Fig. 5 illustrates the fourth step taken toward reaching the exemplary final configuration shown in Fig. 6. The existing dielectric material 12 (shown in Fig. 4) is removed and a new dielectric layer 120 is deposited over both gate electrodes 208, 218 as shown over the nMOS transistor 20. This layer 120 serves to isolate stresses from later-applied layers from reaching the substrate. A blockout photo-resist 22 is applied and the dielectric and stressed film are largely removed around the pMOS transistor 21 by etching. As a result of this step, the nMOS transistor 20 is still influenced by the tensile film 11 and still maintains an enhanced level of performance, while the pMOS transistor 21 is no longer suffering a degradation of performance but resumes a normal potential performance as a result of the removal of the tensile stress provided by the stressed film 11.

Fig. 6 illustrates the fifth and final step in forming the exemplary final configuration as shown. In this step, the photo-resist 22 is first removed from the nMOS transistor 20 side of the chip and then another stressed film 13 is applied across the entire chip. This film preferably provides compressive stress to the pMOS transistor 21 thereby increasing its carrier mobility while having little, if any effect on forces in the channel of nMOS transistor 20. The nMOS transistor 20, as it is separated from the compressive film 13 by a dielectric layer 120, is largely unaffected by this step and remains under the influence of the tensile film 11 only. Now both the pMOS channel 211 and the nMOS channel 211 have enhanced carrier mobility as a

result of this particular application of stressed films in relation to the transistors.

Fig. 7 is a graphical representation of the resulting lateral stress levels applied to the transistors through application of this invention in simulation. This graph is based on the stress level on the transistors at a location 5 nm below the gate dielectrics 205, 215 and a stressed film thickness of 50 nm using stressed nitride for both the tensile film and the compressive film. The resulting stress on the defined areas based on the defined conditions is approximately +190MPa in the nMOS transistor channel 201 and -300MPa in the pMOS transistor channel 211. Thus, it is seen that tensile and compressive stresses are produced substantially coextensive in the transistor channels 201, 211 on the same substrate 14.

An alternate, but similarly effective method for manufacture of transistors in accordance with the invention would involve the same progression of step one (Fig. 2) and step two (Fig. 3) but then implementing alternate steps three, four, and five as illustrated in Figs. 8a, 8b, and 8c, respectively. In Fig. 8a a photo-resist is applied to the nMOS transistor side of the CMOS, thereby protecting it from film removal, retaining a layer of tensile film around the entire nMOS transistor gate electrode 208 and channel 201. The pMOS transistor, not covered by the photo-resist, has the layers of oxide and tensile film removed, as illustrated in Fig. 8b. Following the film removal, the photo-resist is also removed, also illustrated in Fig. 8b. A compressive film is then applied over both nMOS and pMOS transistors to achieve a second

embodiment of the invention as illustrated in Fig. 8c.

The preferred embodiment and second embodiment provide the same function although they exhibit a few small differences in form. The preferred embodiment (Fig. 6) provides a dielectric layer separating the tensile film 11 from the compressive film 13 whereas the second embodiment (Fig. 8c) does not have this layer between films over the gate electrode 208 (although it does exist on the gate sides). Additionally, the tensile film does not fully surround the nMOS transistor 20 in the preferred embodiment (Fig. 6) as it does in the second embodiment (Fig. 8c), but rather contacts the tensile film to a portion of the gate sides only and has a neutral dielectric 120 contacting the rest of the gate electrode 208. This may result in minor dimensional variation between the first and second embodiments depending on the thicknesses of the various films. Also the contiguous compressive and tensile films 13 and 11 over the gate counteract each other in the region of the gate electrode 208.

Further, in regard to the pMOS transistor, the preferred embodiment (Fig. 6) includes a dielectric layer 120,17 surrounding the majority of the oxide liner 19 of the gate electrode 218 with exception to the top of the gate which is left to make contact with the compressive film 13. The second embodiment does not include the aforementioned dielectric layer and therefore, the entirety of the oxide liner 19 surrounding the sides of the gate electrode 218 is in contact with the compressive film 13 which makes direct contact with the top of the gate electrode 218. These differences between the first and second

embodiments do not effect the stress levels within the channels 201, 211 and therefore said first and second embodiments, when compared, exhibit no significant difference from one another in potential carrier mobility.

In view of the foregoing, it is seen that the invention provides a method and structure for controlling or improving the carrier mobility in both nMOS and pMOS transistors on the same chip without compromise of manufacturing yield or adverse effects on previously formed structures or later performed processes and which can be readily controlled without compromising manufacturing yield and requiring only a few additional but well-developed processes. Since compressive and tensile forces, although potentially substantial, are applied over relatively small respective areas (in comparison, for example, to chip thickness) there is no tendency of the chip or wafer to warp or curl. It should be appreciated that while an "improvement" in carrier mobility may generally connote an increase therein, a reduction in carrier mobility may be desired and provided by the same process merely by exchanging the etch-stop layer deposition processes relative to the transistor types to thus reverse the types of tensile or compressive forces are applied to respective transistor conduction/impurity types. Since the amount of tensile or compressive stress can be controlled by deposition process parameters, hole and electron carrier mobility may thus be enhanced, reduced or regulated to any desired degree within the range of the effect of tensile and compressive force thereon depending also on the geometry of the gate.

While the invention has been described in terms
of a single preferred embodiment, those skilled in
the art will recognize that the invention can be
practiced with modification within the spirit and
5 scope of the appended claims.